

UF2216: Repositorios de contenidos

Elaborado por: Dionisio Acosta Álvarez

Edición: 5.1

EDITORIAL ELEARNING S.L.

ISBN: 978-84-16360-87-1

No está permitida la reproducción total o parcial de esta obra bajo cualquiera de sus formas gráficas o audiovisuales sin la autorización previa y por escrito de los titulares del depósito legal.

Impreso en España - Printed in Spain

Presentación

Identificación de la Unidad Formativa

Bienvenido a la Unidad Formativa **UF2216: Repositorios de contenidos**. Esta Unidad formativa pertenece al Módulo Formativo **MF0967_3: Creación y gestión de repositorios de contenidos** que forma parte del Certificado de Profesionalidad **IFCD0211: Sistemas de gestión de información**, de la familia de **Informática y Comunicaciones**.

Presentación de los contenidos

La finalidad de esta unidad formativa es enseñar al alumno a crear componentes software para almacenar de modo estructurado la información de una organización, según el diseño especificado y estándares definidos por organismos de normalización.

Para ello, se analizarán los repositorios de contenidos, la estructura interna de un repositorio y los metadatos. Por último, se estudiarán los sistemas de gestión de contenidos y los sistemas de gestión del aprendizaje (LMS).

Objetivos de la Unidad Formativa

Al finalizar esta Unidad Formativa aprenderás a:

- Crear repositorios de contenidos utilizando lenguajes específicos y estándares de desarrollo software.

Índice

UD1. Repositorios de contenidos

1.1. Conceptos básicos	11
1.2. Incorporación de contenidos.....	20
1.2.1. Adquisición	21
1.2.2. Agregación.....	23
1.2.3. Creación	25
1.3. Asociaciones entre contenidos	26
1.3.1. Jerarquías, taxonomías	28
1.3.2. Índices	29
1.3.3. Referencias cruzadas	35
1.4. Ciclos de vida de los documentos	37
1.5. Flujos de trabajo.....	40
1.6. Versionado de contenidos	42
1.7. Check In/Check Out.....	47
1.7.1. Resolución de conflictos	49
1.7.2. Herramientas diff/merge	50
1.8. Búsqueda y localización de la información.....	51

UF2216: Repositorios de contenidos

1.9. Publicación	52
1.9.1. Sindicación	55
1.9.2. Servicios Web	57
1.10. Seguridad	63

UD2. Estructura interna de un repositorio

2.1. Almacenamiento de la información	81
2.1.1. Base de datos.....	94
2.1.2. Ficheros	111
2.1.3. Bases de Datos y ficheros	113
2.2. Estándares de desarrollo	126
2.2.1. CMIS	127
2.2.2. JSR-168	133
2.2.3. JSR-170	140
2.2.4. JSR-283	149

UD3. Metadatos

3.1. Concepto	167
3.2. Beneficios	172
3.3. Ontología RDF.....	175
3.4. Metadatos Dublin Core.....	189
3.5. Metadatos XMP (Adobe's Extensible Metadata Platform)	194
3.6. Metadatos EXIF (Exchangeable image file format).....	199
3.7. Metadatos IPTC	202

UD4. Sistemas de gestión de contenidos

4.1. Sistemas de gestión de bitácoras (Blogs)	217
4.2. Sistemas Wiki	260
4.3. Sistemas Portales.....	319

UD5. Sistemas de gestión de aprendizaje

5.1. Descripción General	335
--------------------------------	-----

Índice

5.2. Estándares IMS	342
5.3. Estándar IEEE LOM	359
5.4. Iniciativa ADL SCORM	368
Glosario	381
Soluciones	383
Anexo	385

Área: informática y comunicaciones

UD1

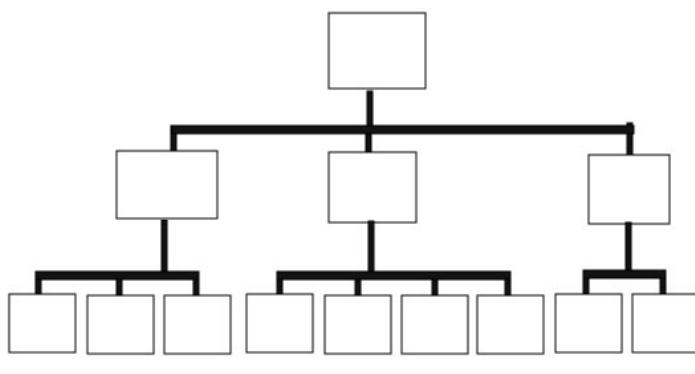
Respositorios de contenidos

UF2216: Repositorios de contenidos

- 1.1. Conceptos básicos
- 1.2. Incorporación de contenidos
 - 1.2.1. Adquisición
 - 1.2.2. Agregación
 - 1.2.3. Creación
- 1.3. Asociaciones entre contenidos
 - 1.3.1. Jerarquías, taxonomías
 - 1.3.2. Índices
 - 1.3.3. Referencias cruzadas
- 1.4. Ciclos de vida de los documentos
- 1.5. Flujos de trabajo
- 1.6. Versionado de contenidos
- 1.7. Check In/Check Out
 - 1.7.1. Resolución de conflictos
 - 1.7.2. Herramientas diff/merge
- 1.8. Búsqueda y localización de la información
- 1.9. Publicación
 - 1.9.1. Sindicación
 - 1.9.2. Servicios Web
- 1.10. Seguridad

1.1. Conceptos básicos

Podemos definir un repositorio de contenido como aquel lugar, depósito o archivo/s donde se encuentra almacenada y centralizada la información de diversa índole en ficheros informáticos y en bases de datos.



Estructura de Almacenamiento

Un repositorio de contenidos es muy parecido a un servidor de archivos, con la diferencia que el repositorio puede recordar las modificaciones que sobre los archivos o ficheros se hayan realizado así como los cambios en el directorio, estos cambios pueden ser:

- Incorporación o adición de nuevos ficheros.
- Borrado de ficheros.
- Reubicación de ficheros y directorios.

Frecuentemente encontraremos repositorios de contenidos alojados en servidores web dentro de internet, sin embargo también existen repositorios de conte-

nidos que están alojados en medios de almacenamiento extraíbles o portátiles como discos duros. Lo fundamental en ambos casos, y para que hablemos con propiedad de un Repositorio de contenidos en toda regla es que este cumpla con el concepto indicado en el Revision Control System – **RCS** o **Sistema de Control de Versiones**, que indica que lo realmente importante es que se pueda realizar un seguimiento y control de los cambios generados en cada uno de los ficheros de ese repositorio o proyecto. Debe contener un historial de versiones de todos y cada uno de los elementos que en él se han gestionado.

Un **control de versiones** debe cumplir con las siguientes características:

- Almacenamiento de ficheros de gestión en cualquier formato, bien sean documentos de texto, ficheros multimedia, fotografías, archivos digitales, etc.
- Indicar las modificaciones parciales o totales que han sido realizadas en los elementos contenidos en el CMS. Ficheros borrados, eliminados, movidos o renombrados.
- Debe ofrecer un Registro histórico de las modificaciones o acciones causadas a cada elemento y a su conjunto, incluyendo la opción en la que podremos volver a una o varias versiones anteriores de cada elemento.

Así, los repositorios de contenidos permiten la publicación de diversa información de una manera fácil y rápida, estructurada, jerarquizada y permite:

- Ser separado según el tipo de contenido o formato: inicialmente el contenido no debe contener ningún tratamiento gráfico. La aplicación de tratamiento gráfico dará diferentes opciones de presentación (imágenes, video, galería fotográfica, presentaciones multimedia, etc.) que pueden ser almacenadas según el tipo de fichero, el tipo de usuario que accede a este repositorio, o según el dispositivo desde donde se quiera acceder.
- El material bruto con el cual se generará el contenido será administrado separadamente del material final generado (contenido desarrollado); por un lado estarán almacenados los textos, videos, sonidos, imágenes y fuentes a utilizar y por otro el material resultante, el contenido generado.
- Ser accesible desde una interfaz amigable para los usuarios.
- Modificación de los ficheros o documentos que se encuentran dentro del ciclo de publicación establecido por el Administrador del Repositorio, evitando así que un mismo documento sea editado por un usuario mientras que es modificado por otro.

- Otorgar permisos de acceso: para cada tipo de usuario habrá determinados permisos de acceso, modificación y control de la información, este acceso es completamente personalizable.

Un Repositorio de Contenidos o Sistema de gestión de contenidos, en adelante CMS por sus siglas en inglés Content Management System, es:

Un Sistema que permite la gestión accesible, rápida, cómoda y uniforme de uno o varios clientes o personas a una base de datos actualizada periódicamente, con información o contenido relevante para un interés general. Este sistema se apoya en bases de datos y está conformado por un conjunto de programas incluidos en un servidor web que con otro conjunto de programas o aplicaciones de acceso para clientes permitirán el acceso a esa información.

Un Repositorio de contenidos debe ofrecer las herramientas que proveerán el proceso de creación, materialización o realización y presentación del contenido a los usuarios del sistema, y por otro lado debe ofrecer mantenimiento y actualización de los ficheros y/o archivos que en él se encuentran así como de su contenido.

Por ello en un Repositorio de contenidos encontramos dos tipos de relaciones:



- Con los usuarios del sistema: para los que habrá que realizar una gestión inmediata, accesible, amigable y uniforme al usuario del sistema, ya que todos buscan información de diverso tipo y formato pero con una característica común: actualizada y actualizable periódicamente. La funcionalidad y practicidad, así como la sencillez y usabilidad son características fundamentales.
- Con los desarrolladores: la idea es crear un repositorio de contenidos online orientado siempre al usuario y para ello los desarrolladores deben cumplir con las normas y estándares aceptados globalmente en cuanto a apariencia, interfaz, usabilidad, etc. Deben facilitar la creación y presentación de los contenidos, facilitando su publicación en varios formatos, y permitiendo su categorización y navegación.



Relaciones

Podríamos realizar una clasificación de los Repositorios de contenidos de la siguiente manera:

Telemáticos

Institucionales

Para Software

- **Repositorios según temas o temáticos:** este tipo de repositorios deben proveer a los usuarios acceso a un contenido específico, particular. Podremos incluir en este tipo aquellos repositorios que tienen un carácter pedagógico o didáctico, deben incluir información sobre objetos y metodología del aprendizaje, ordenada cronológicamente y clasificada mediante etiquetas o metadatos que faciliten la búsqueda del contenido. Este temática suele ser de tipo social, pueden ser repositorios temáticos aquellos de carácter: científico, médico, farmacéutico, de diseño gráfico o web, moda, belleza, etc.
- **Repositorios institucionales:** suelen ser desarrollados por instituciones tales como universidades, colegios profesionales, asociaciones y/o organismos políticos y contienen generalmente tesis doctorales o trabajos de investigación y revistas o publicaciones. El contenido de este tipo de repositorios suele venir definido por la institución, por lo general tienen carácter instructivo o académico, son perpetuos y acumulativos, abiertos y poseen capacidad para funcionar con otros sistemas ya existentes o futuros (interoperabilidad). Este tipo de repositorios puede variar en tamaño, desde pequeños repositorios como los que encontramos en colegios profesionales o algún centro de investigación local a grandes como los ofrecidos por organizaciones internacionales.
- **Repositorios de/para software:** permiten el acceso a multitud de aplicaciones que podrán ser descargadas e instaladas rápida y cómodamente. Nos permiten la selección de aquellas herramientas que sean de nuestro interés y según el tipo de licencia podremos encontrar:
 - **Repositorios de Licencia Privada:** donde el software está limitado o restringido a las condiciones que impone el fabricante o desarrollador. Windows Update podría ser un ejemplo de repositorios de licencia privada.
 - **Repositorios de Licencia de uso libre:** de uso cada vez mayor, el software libre permite al usuario que lo adquiere (gratuitamente o mediante pago) la ejecución, copia, distribución del software. Pueden ser ejemplos de Repositorios de Licencia de software de uso libre: SourceForge y Forja, que ofrecen paquetes para GNU/Linux, y ofrecen plataformas en donde el trabajo colaborativo y compartido del conocimiento es la base, no existen restricciones de licencia y se puede acceder a listas de email, foros, entre otras funcionalidades.

En cualquiera de los casos anteriormente mencionados, se deben cumplir, entre otras, con las siguientes características:

Los documentos deben ser localizables permanentemente mediante identificadores	Deben ser aceptados y utilizados según el protocolo los estándares de metadatos, descriptivos, conservación y administrativos
Deben ser interoperables (norma OAI-PMH, Z39.50)	Deben apoyarse en formatos de archivo, escalabilidad, extensibilidad y mantenimiento de los sistemas
Deben contener campos o aplicaciones de búsqueda y visualización de metadatos	

Según lo que hemos visto en este punto, un repositorio de contenido es un medio donde se almacena contenido digital, para su acceso mediante opciones de búsqueda y recuperación de la información, que vendrán dadas mediante la implantación de instrumentos que hagan posible la importación, identificación, almacenamiento, preservación, muestra y exportación de esa información digital. Con todo ello es de destacar la importancia de los repositorios de contenido para la comunidad, la investigación e investigadores, y las instituciones.

A estas ventajas comunes para todos los tipos de repositorios, habrá que añadir aquellas que ofrecen los repositorios de acceso abierto, como es la de ofrecer una mayor comunicación y difusión de las obras intelectuales, técnicas, artísticas o científicas sin restricciones de registro, suscripción o pago por el acceso a estos contenidos.

Ventajas de los repositorios de contenidos:

- Amplían información gracias al libre intercambio de la información.
- Proveen de conocimiento a organizaciones, grupos y a la comunidad en general. De esta forma promueven la reutilización y compartición del contenido y por ende del conocimiento.
- Potencian la divulgación de los procesos de investigación.
- Interoperabilidad con otros sistemas propios de la institución y otros repositorios de contenidos.
- Fomentan la especialización al otorgar contenido de valor.
- Estandarización de los contenidos.
- Proporcionan información veraz y actualizada en cuanto a novedades en el sector.

Existe una serie de factores que propiciarán que un repositorio de contenidos tenga la eficiencia deseada, a saber:



Objetivos del Repositorio de contenidos:

1. **Recopilar y organizar:** los contenidos digitales desde su creación, así como de todos aquellos datos que los describirán clara y correctamente, con un lenguaje estandarizado. Con esto se logrará que una vez almacenados, los usuarios puedan encontrarlos fácilmente.
2. **Visibilidad y acceso:** teniendo como premisas los derechos de autor, la visibilidad de la información, y la seguridad de los contenidos, estos deberán ser accesibles a los usuarios que los busquen proporcionando la información completa y a los sistemas informáticos mediante datos descriptivos.
3. **Preservación y mantenimiento:** tomando en cuenta que el contenido puede sufrir transformaciones o ser objeto de migraciones a otros repositorios o software, y que este contenido debe ser mantenido durante largos periodos de tiempo, debemos proveernos de herramientas adecuadas y necesarias para que no se produzca pérdida de ficheros e información, garantizando así la seguridad de la información, los documentos o ficheros y la del repositorio en sí mismo. Este punto debe mantenerse desde el inicio mismo del repositorio, ya que la instalación de un software que permite la gestión del repositorio debe asegurar: efectividad, permanencia e interoperabilidad con los sistemas internos de la empresa o institución y con otros sistemas externos.

En cuanto a la Interoperabilidad del Repositorio debemos considerar diversos aspectos:



En relación a la organización:

- Debe proveerse de una metodología correcta de recolección de información y contar con procesos, preferiblemente de auto almacenamiento o auto depósito por parte de los usuarios.
- Verificación de los derechos de autor, para cada uno de los documentos que serán recopilados, para garantizar el control de estos y del repositorio en general.
- Definición en detalle de roles, funciones, flujos y procesos de trabajo, así como la integración de estos a los manuales operacionales de la organización. A ser posible debe también aportarse información sobre el desarrollo de habilidades y competencias de los distintos participantes del repositorio.

En relación a los contenidos:

- Deberán estar ordenados según a la tipología de cada contenido, y preferiblemente en base a lo establecido en directrices internas, de carácter nacional o internacional ya existentes.
- Garantizar que los documentos publicados cumplen completamente las directrices establecidas por el autor o empresa, institución en cuanto a sus derechos.
- Definir la catalogación de los documentos, manejo de archivos y metadatos asignados a cada contenido.
- Garantizar la preservación de los contenidos, definiendo los procesos necesarios para este fin.

En relación a la tecnología:

Planificar detalladamente las necesidades o requerimientos tecnológicos del repositorio.

- Seleccionar el sistema más idóneo según las necesidades que a largo plazo pueda presentar la organización.

En relación a la normativa:

- Utilización de vocabulario aceptado según los estándares internaciones para descriptores de la información tales como: autores y áreas temáticas, idiomas y/o países, formatos de archivos, fechas de creación, publicación, etc.

- Usar los identificadores únicos internacionales en los documentos.
- Según el tipo de organización se podrá estudiar el apego a las normas de intercomunicación internacional para repositorios de contenidos.

En resumen, para que exista un repositorio de contenidos propiamente dicho deben darse las siguientes características.

- Acceso mediante ordenadores conectados al contenido del Archivo Digital.
- Utilización de redes de comunicación.
- Debe contener documentación producida y recibida por una institución o persona.
- La documentación debe ser un proceso que consta de los siguientes pasos:



- El contenido deber ser recuperado mediante búsquedas por usuarios o sistemas.
- El contenido debe estar sujeto a criterios de descripción mediante el uso de:
 - Normas.
 - Cuadros de clasificación.
 - Etiquetas.
 - Descriptores.
 - Lenguajes controlados.
 - Campos.
 - Metadatos.

- El objetivo del repositorio es difundir y dar a conocer los fondos documentales y el contenido del mismo, para suplir las necesidades de información de los usuarios.

1.2. Incorporación de contenidos

La incorporación de contenidos a un repositorio debe hacerse de la forma más práctica y fácil de realizar posible, sin perder por ello la exactitud, eficiencia y funcionalidad del repositorio. Esta práctica debe incluir entre otros aspectos:

Captación del contenido

Comprobación de derechos de autor

Autorizaciones

La carga o captación del contenido empieza en el momento justo en que se pone en funcionamiento el repositorio, y se va sucediendo indefinidamente puesto que siempre se podrá agregar más contenido catalogado a ese repositorio.

La serie de pasos ordenados a seguir para asegurar que el contenido incorporado al repositorio pueda llegar a los usuarios de este, se debe realizar mediante la definición de un flujo de trabajo o workflow. Este concepto no es específico de la informática pero sí constituye una parte fundamental para el trabajo colaborativo. Mediante aplicaciones de flujos de trabajo se relacionarán y automatizarán secuencialmente las tareas, acciones y actividades para la ejecución de un proceso, incluyendo la opción de poder efectuar en todo momento del proceso un seguimiento de estado.

La finalidad de un sistema de flujo de trabajo en el trabajo colaborativo es la de minimizar el tiempo de ejecución de las tareas, acercando recursos y herramientas a los procesos, permitiendo de esta manera que diferentes personas en diferentes equipos y diferentes lugares puedan realizar una tarea grupalmente.

En caso de que el contenido que queramos incorporar al repositorio sea de dominio público y tenga derechos de libre distribución, no nos encontraremos con ningún problema, podremos dar opciones de acceso y descarga libre a los usuarios internos o externos que requieran este contenido.

El problema se presenta cuando el propietario, presenta o registra su obra con derechos sobre el contenido, derechos que en el medio digital vienen amparados por una serie de licencias Creative Commons, que aseguran al autor o propietario del contenido sus derechos.